



# Fifteen shadows of socio-cultural AI: A systematic review and future perspectives

Katalin Feher<sup>a,b,\*</sup>, Attila I. Katona<sup>c</sup>

<sup>a</sup> Senior Research Fellow, Centre of Excellence on Cybereconomy, Budapest Business School University of Applied Sciences, Budapest, Hungary

<sup>b</sup> Fulbright Research Fellow, Department of Communication, Drexel University, Philadelphia, United States

<sup>c</sup> Research Fellow, Department of Quantitative Methods, University of Pannonia, Veszprem, Hungary

## ARTICLE INFO

### Keywords:

Artificial intelligence  
Socio-cultural AI  
Techno-AI  
Policy making  
Systematic review  
Topic modeling

## ABSTRACT

The number of studies related to socio-cultural AI (SCAI) is growing dramatically. Therefore, the goal is to perform the first systematic review of the key sources published over the last decade with consequences for social science, humanities, engineering, computer science, and policy research. The novelty of the study is not only the first snapshot of high-ranked articles from seven academic databases but also the revealed and interpreted SCAI research trends with implications for academia and policymaking. Topic modelling is conducted on 607 papers identifying fifteen well-defined fields. Association networks also unfolded trending research areas with smart cities, cultural-creative industries and media. A timeline of the emerging research topics reveals the year of change for SCAI was 2018, mostly with industry 4.0, governing AI, and smart cities. Last but not least, SCAI research for policies is interpreted as a niche for policymaking and academic research funding. The findings summarize the broad coverage of AI technology in society and culture with related research responsibility as underrepresented topics. Implications and weak yet relevant signals are also formulated for academic and policy research. The main contribution of this study is to discover the SCAI research for academic research and policymaking for future perspectives.

## 1. Introduction

This study distinguishes two approaches to artificial intelligence, namely techno-AI and socio-cultural AI. Both approaches are suggested as categories in this paper to understand the inevitable embeddedness of state-of-the-art technology in society and culture.

The first proposed term is techno-AI. There is no available a definition of techno-AI so far but it is suggested to distinguish the technology defined AI from the socio-cultural aspects of AI. Techno-AI is a category for computing based technology producing AI developments and focusing on STEM-fields (science, technology, engineering, and mathematics) behind the innovations from robotic automation to machine learning. This approach emphasizes the AI-technology empowering in several industries. Therefore, AI-technology and its developments are the final goal in themselves to the competitive advantage. According to the leading academic databases, such as Web of Science and Scopus, techno-AI research publications are representing mostly computer science, engineering, and scientific fields have gathered momentum over the past years.

\* Corresponding author at: Senior Research Fellow, Centre of Excellence on Cybereconomy, Budapest Business School University of Applied Sciences, Budapest, Hungary.

E-mail address: [feher.katalin@uni-bge.hu](mailto:feher.katalin@uni-bge.hu) (K. Feher).

<https://doi.org/10.1016/j.futures.2021.102817>

Received 20 October 2020; Received in revised form 2 July 2021; Accepted 12 July 2021

Available online 24 July 2021

0016-3287/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The second proposed term is the socio-cultural AI (hereinafter SCAI) focusing on the technology as a tool and not as an ultimate goal. Social science and humanities define the AI-related socio-cultural issues with wide range of topics from AI-related policy making to fake news. The SCAI category describes the embeddedness of AI-technology in the fabric of society and culture from human adaptation to ethical dilemmas. Although the broad category of “society and culture” assumes numerous academic approaches and research options, the leading databases present a suitable number of high-ranked publications for analysis in relation to technology. According to Scopus analytics, the number of socio-cultural publications in technology, and also, in AI has jumped in the last ten years and mostly from 2018. The key subject areas are social science, computer technology, engineering humanities, and medicine. The noticeable decade-long period and the proportion of the subjects confirm the relevance of the broad and synthesized term of SCAI. The SCAI-approach is also confirmed by the related and growing interest in public or business policy, and also the cultural norms or AI ethics, thus in socio-cultural technology (Coeckelbergh, 2020; Diallo, Shults, & Wildman, 2021; Rizzo, 2020).

If we accept that society and culture are both drivers and inhibitors of technological acceleration, the current question is how techno AI and SCAI are currently growing relative to each other. The expected research trends of SCAI are rooted in this distinction. The number of publications is increasing noticeably in both categories. Further subjects or research methodologies are expected in the near future along with the emerging AI extensions. However, when the academic research outputs of both categories are compared, techno-AI is highly represented, while SCAI is underrepresented in the same databases.

A comprehensive summary investigating the academic publications in SCAI-research has not been available yet. However, the emerging fields of SCAI allow the first summary and future forecast of this field. The purpose of this paper is to reveal and classify the key subjects in this field, in particular reference to emerging and expected research. This paper uses topic modeling, trend analysis, and association networks of qualified academic research to systematize the discourse of SCAI. An overall view of the SCAI literature supports government and business decisions like funding, innovation and start-up projects, regulation and sales of techno-AI. The results of this study provide access to relevant and studied topics for those considering options for innovation. Additionally, the paper supports understanding the fabric of society and cultural inclusion in the case of techno-AI developments and their future options. The field of “SCAI” is discovered to implications and future perspectives in academic and policy research.

The rest of the paper is organized as follows: Section 2 outlines theoretical considerations, Section 3 presents the research goal and methods, Section 4 shows findings, Section 5 describes the analyses related to validation, and Section 6 offers a conclusion, discussion and recommendation to the future perspectives.

## 2. Theoretical considerations

There is a complex relationship between advanced digital technology and the socio-cultural environment. It is particularly essential to systematize the key topics for diverse subjects ranging from viral cartography (Robinson, 2019) or climatic conditions (Kaur & Bala, 2018) to fair and responsible autonomous systems (Brand ao, Jirtoka, Webb, & Luff, 2020) or social media and psychoanalysis (Emmert-Streib, Yli-Harja, & Dehmer, 2019), to mention only a few examples from a large body of SCAI research. Certain authors propose a wide scope of SCAI research, representing multidisciplinary study or philosophical approaches, such as Dwivedy's (Dwivedi et al., 2019) agenda for research and policy, Lee's (Lee & Cook, 2020) myth of the data-driven society, Makridakis's (Makridakis, 2017) question of utopian or dystopian future, or Galanos (Galanos, 2018), who denies the existence of AI in relation to natural intelligence. This kind of summary usually focuses on human-centric, ontological, or ethical questions in a complex way. However, these are not systematic literature reviews of SCAI. Research, business, and policymaking require a more comprehensive summary than this kind of research offers. Therefore, the purpose of this paper is to discover the key research trends of SCAI through the first systematic literature review.

The tao model of technology proposed by Seel (2012) seems to be the most useful theory for analyzing the relationship between techno-AI and SCAI. According to his yin-yang model, yang is the active, light part of the dual tao model, representing vibrant technology, while yin is the dark and passive part that emphasizes the social impacts of technology. The yin-yang model would be more precise if it represented society not as passive but as a dynamic entity that simply changes more slowly than technology does. Additionally, this model should be expanded to include aspects of particular cultures; such an expansion would enable value-oriented and contextual interpretations of the relationship between technology and society (Feher & Zelenkauskaitė, 2020; Felzmann, Vilaronga, Lutz, & Tamò-Larrieux, 2019). This way, the two elements of the yin-yang model are understood to be both active and reactive and to influence each other dynamically. This influence is especially noticeable when a technology starts to spread rapidly, reaching a tipping point (Kekäle & Helo, 2014; Phillips, 2007), facilitating changes in society and culture. These trends present significant patterns, and also, weak signals (Lee & Park, 2018) of the future.

Although the papers cited above investigate the relationship between technology and society, they are not systematic in nature. Therefore, the socio-cultural embeddedness of AI has not been explored with a sufficient degree of systematicity. Consequently, there is a fundamental question: what are the leading topics of SCAI research to influence the future AI-adaptions? In other words, which topics can be identified clearly as describing either the impact of technology on socio-cultural concerns (Rubin, 2019) or the impact of socio-cultural values on technology (Sunny, Patrick, & Rob, 2019)?

In summary, techno-AI is facilitated and evaluated by society and culture. Research of SCAI is emerging slowly, as the yin-yang model emphasizes. Thousands of techno-AI studies are published every year, while SCAI appears only in hundreds of records in leading academic databases. However, a sufficient number of publications are already available to enable researchers to classify which technological implementations are increasingly convergent with society and culture. This study and its findings provide a comprehensive summary of this field with future forecasts.

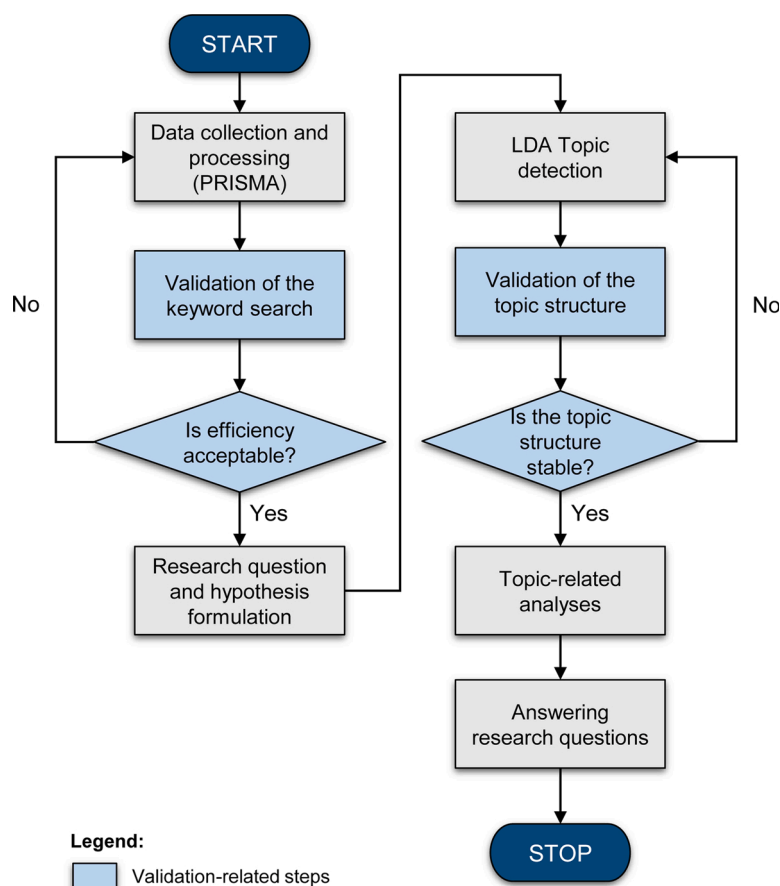


Fig. 1. Framework of the analysis.

### 3. Research goal and methods

#### 3.1. Structure of the review

In this paper, we applied several steps from the data collection to the synthesis of research findings. Fig. 1 shows the flowchart related to the structure of data collection and analysis. Figure shows the flowchart related to the structure of data collection and analysis.

During the data collection and data processing, the PRISMA statement is applied including its suggested phases such as *identification, screening, eligibility and included* (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009) (see detailed description in Section 3.3). After this step, we investigate the efficiency and robustness of the keyword search by measuring the relative number of relevant articles when additional keywords are added to the search (Section 3.5). If additional keywords cannot significantly increase the set of relevant papers, the initial keyword set is considered to be valid. After the formulation of research questions and hypotheses, a topic modeling approach is used to reveal the main topics of SCAI (Section 3.4). The extracted topics are also validated using an additional, community detection-based method (Section 3.5). If the two approaches lead to similar results, the extracted topic labels are considered robust enough to continue the analysis without parameter tuning. Finally, we conduct topic-related analyses to answer the research questions (Section 4).

#### 3.2. Goal and research questions

The goal is to discover research trends of SCAI through the first systematic review from seven academic databases published over the last decade with implications for social science, humanities, computer science, and policy research. Accordingly, research questions were formulated as follows:

- RQ1. What are the leading topics of SCAI in academic research?
- RQ2. Which SCAI research topics are the most influential and which are on the periphery with less focus?
  - RQ2.1 What are the core converging SCAI research topics?

–RQ2.2 Which SCAI research topics are not directly interconnected and are on the periphery?

- RQ3. What is the hierarchical relationship between the leading SCAI topics?
- RQ4. What have been the trending SCAI topics over the past decade with expected future expansion?
- RQ5. What are the SCAI research trends for policies?

A comprehensive summary of the answers to these research questions is available below.

### 3.3. Systematic review and data collection

A systematic literature review (SLR) is a methodologically-rigorous synthesis of the highly-qualified scientific literature on a specific topic or research question. Its goal is not just to aggregate the results of research on the selected topic but also to support the construction of evidence-based research guidelines (Kitchenham et al., 2009). To produce an SLR, this study used the PRISMA methodology, which was proposed by Moher et al. (2009) for the analysis of bodies of literature. All four stages proposed by the PRISMA method, (1) “identification,” (2) “screening,” (3) “eligibility,” and (4) “included,” were employed by this study.

**(1) Identification.** First, we tested the quantity and quality of the possible results. Thus, time range, document type, and language needed to be specified. The time span was defined as the above-mentioned Scopus analytics of technology-related socio-cultural research and their high-ranked publications, and also, as the uplifting techno-AI developments from deep learning to personal assistants. Thus, the time range is one decade from 2009 to 2019. The corpus came from abstracts and keywords of the academic sources because this option is available in all key databases and the selection assumes a strong focus on SCAI-related fields. The filtered language of the abstracts and keywords was English for the largest number of sources without meaning anomalies. Only the abstracts and keywords were limited to English to maximize the sample size by allowing multiple languages.

Considering the AI definitions by the leading dictionaries (Cambridge, Oxford Reference, Merriam-Webster, Collins, and Mac-Millan), “digital technology” and “computing technology” were the part of the first testing keywords beyond “artificial intelligence”.

Considering the extended yin-yang model above, keywords were “society” and “culture” as their versions with nouns and adjectives. Scopus, as the biggest relevant database, was tested first with these keywords. There were almost twenty thousand search hits with the above-mentioned filters and the searching rule:

*(technolog\* OR digital OR comput\*) AND (social OR society OR socio) AND (culture OR cultural)*

Filtering the technology part of the rule and focusing only on “artificial intelligence” with the same extended filtering method, the rule was

*(“artificial intelligence”) AND (social OR society OR socio) AND (culture OR cultural)*

resulting in 239 hits. These tests present a large number of articles in digital and computing technology in a socio-cultural context while the AI-related fields are only one percent in this proportion. This result refers to a specific field that is already suitable for analysis. However, an expansion with a further relevant database was necessary for quantitative research. The search rule with “artificial intelligence” was applied for the further six databases as their specific search systems.

Related technologies were also tested from machine learning to robotics, but these were specific and less broad topics within digital and computing technology compared to AI. The cultural aspects were underrepresented compared to the social factors in the studied databases. Thus, “socio-technical systems” or similar options were excluded to keep studies that also had a cultural focus.

After database testing, benchmarking also supported the research decisions. AI is a well-defined technology with a system to act intelligently (Di Vaio, Palladino, Hassan, & Escobar, 2020) or with cognitive, emotional and social intelligence (Haenlein & Kaplan, 2019) from economy or politics to morality or art (Vanderburg, 1985). In this sense, AI requires special attention among all the technologies as it provides a mimicry of human intelligence and starts to act like humans (Bloom, 2020), thus comprehensively affecting social and cultural functions. These functions were investigated together in this paper with their overlapping fields (see Section 4).

Beyond Scopus and for more high-ranked academic hits, six additional relevant databases were studied together, these were EBSCO Information Services, JSTOR Digital Library, ArXiv e-print Service, Directory of Open Access Journals, Social Science Research Network, and Institute of Electrical and Electronics Engineers. There is no available similar selection for a comprehensive study in social or cultural related AI. Following recent scientific trends (Tight, 2018), only journal articles were selected for high-ranked academic outputs. Although these databases apply different search algorithms and they use various and changing indexing methods, the goal was to produce diverse and multiple data collections to study the academic sources broadly. Obviously, this research ambition assumes uncertainties as opposed to research relying on a single database. However, the emerging and complex field required comprehensive data collection with snapshot research with an acceptable amount of uncertainty. Considering all these, snapshot research is supported to reveal diverse academic publications of SCAI within a decade time range.

**(2) Screening.** A three-step data consolidation process was necessary. The first step was the removal of duplicates. Most of the sources were present in only one database, but more than ten percent of sources were detected two or more times. After the duplicates from the initial 754 abstracts were eliminated, 607 abstracts remained, and 147 were deleted. Second, not all scientific articles present keywords directly. The missing keywords were retrieved from further sources. In 31 cases, the keywords were not available at all. This was an acceptable amount of missing data since it was about five percent of the remaining 607 abstracts. Third, certain abstracts contain copyright information or other metadata. These were deleted from the texts to ensure accurate topic mining.

**(3) Eligibility.** Only the abstracts and keywords of the journal papers were investigated to establish which topics were closely related. This study was not intended to analyze full-length articles that only mention social or cultural aspects but do not have a strong focus on them.

**(4) Included.** Finally, 607 abstracts and 576 branches of keywords were used in this study.

### 3.4. Topic modeling

In order to conduct a comprehensive review of the literature in the studied research field, selected abstracts were first analyzed using topic modeling. The Latent Dirichlet allocation (LDA) was used for topic modeling because it is a widely-used topic mining approach that can reveal the “hidden” structures of a text collection (Chen et al., 2017; Jeong, Yoon, & Lee, 2019; Maier et al., 2018; Papadimitriou, Tamaki, Raghavan, & Vempala, 1998). Three principal concepts are the basis of LDA topic modeling: (1) *corpus*, which denotes the text collection, (2) *documents*, which are the items of the corpus (in the case of this paper, these items are the abstracts) and (3) *terms*, which denotes the words in the documents.

LDA represents documents as a distribution over latent topics, and each topic is represented by a distribution of terms. Let  $D$  denote the number of documents,  $T$  the number of topics, and  $V$  the dictionary containing the unique list of terms from the corpus. Furthermore, let  $N_d$  be the number of terms within  $d^{\text{th}}$  document. The generative process can be described as follows (Jelodar et al., 2019):

- 1 For each topic  $t(t \in \{1, \dots, T\})$ :
  - (a) Choose a word distribution  $\vec{\varphi}_t \sim \text{Dir}(\beta)$
- 2 For each document  $d(d \in \{1, \dots, D\})$ :
  - (a) Choose a topic distribution  $\vec{\theta}_d \sim \text{Dir}(\alpha)$
  - (b) For each word  $w(w \in \{1, \dots, N_d\})$ :
    - i. Select a topic  $z_n$  from Multinomial( $\vec{\theta}_d$ )
    - ii. Select a word  $w_n$  from Multinomial( $\vec{\varphi}_{z_n}$ )

Where  $\text{Dir}(\alpha)$  is a Dirichlet distribution with  $\alpha$  parameter and  $\text{Dir}(\beta)$  denotes Dirichlet distribution with parameter  $\beta$ . Here,  $T$ ,  $\alpha$ , and  $\beta$  are hyperparameters. The assumption of the  $\theta$  and  $\varphi$  parameter estimation is that the number of topics ( $T$ ) is known. In many cases,  $T$  is unknown, and therefore it needs to be estimated using different approaches (Hou-Liu, 2018).

Griffiths and Steyvers (2004) choose  $T^*$  (optimized value of  $T$ ) where the harmonic mean of sampled log-likelihoods (retrieved by Gibbs-sampling) is maximal. Cao, Xia, Li, Zhang, and Tang (2009) estimate that  $T^*$  minimizes the average cosine similarity between topic distributions. Another approach developed by Arun, Suresh, Madhavan, and Murthy (2010) minimizes the symmetric Kullback-Liebler divergence between matrix values, which represents word probabilities related to each topic and the distribution of topics within the analyzed corpus. Although there are several approaches to determine the number of topics, in this study, these three widely-used metrics were applied together to specify  $T^*$ .

### 3.5. Validation of keyword search and the model results

#### 3.5.1. Validation of the keyword search

In this paper we use the term “artificial intelligence” as keyword besides the society and culture-related terms described by Section 3.3. In order to investigate if this rule is efficient enough to identify the relevant papers in the selected area, we conduct a sensitivity analysis containing the following steps:

- 1 A specific scientific platform is selected to conduct the sensitivity analysis.
- 2 The amount of relevant papers ( $r_0$ ) is calculated by applying the initial search rule described by Section 3.3.
- 3 A  $k_i$  list of additional keywords (like “machine learning”, “deep learning”, “computer system”) is defined besides the initial one.
- 4 The rule is extended with an additional keyword from the set defined by step 3 as follows:  
 (“artificial intelligence” OR  $k_i$ ) AND (social OR society OR socio) AND (culture OR cultural)  
 where  $k_i$ ,  $i = 1, 2, \dots, n$  is the  $i$ th keyword from list  $k$ .
- 5 Number of relevant papers found by the modified rule is determined ( $r_i$ ).
- 6 Steps 4–5 are iterated until all of the keywords from  $k$  are used.

Number of relevant papers provided by only the  $i$ th keyword can be calculated as follows:

$$r_i^* = \frac{r_i - r_{i-1}}{r_0} \quad (1)$$

where  $r_i^*$  represents the own contribution of the  $i$ th keyword to the efficiency of the search from the perspective of this paper. Furthermore,  $r_n$  reflects the ratio of additional relevant papers (compared to  $r_0$ ) when all the defined keywords are used. We use this approach for two purposes: (1) to analyze the sensitivity of the search to additional keywords and (2) to validate the initially defined search rule.

#### 3.5.2. Validation of the topic model results

A network community-based solution was used to validate the findings produced using LDA. The network community-based

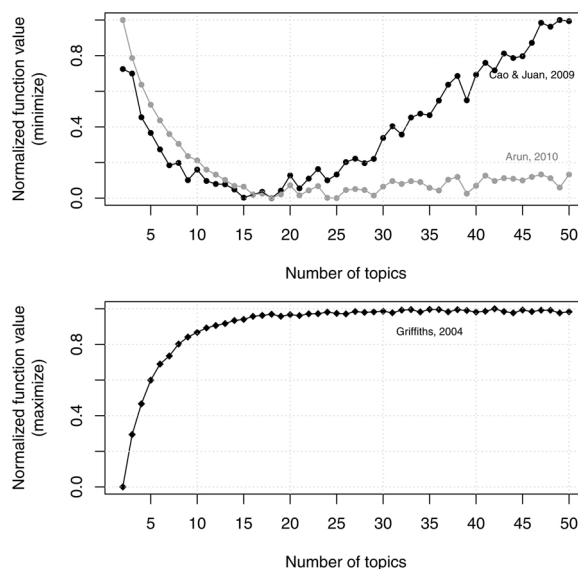


Fig. 2. Optimal topic number selection (simulation).

solution was very different from the LDA topic modeling, so it could verify the findings produced using LDA. The network community-based validated LDA in two ways: (1) validating the number of research topics in the field of SCAI and (2) confirming the key topics based on the abstracts.

For its purpose, the keywords of the scientific papers were also extracted. A network of papers was drawn with this method, where nodes represent the papers and edges indicate that the papers share common keywords. Jaccard's method provided the ability to measure the similarity between the keywords of the papers (Niwattanakul, Singthongchai, Naenudorn, & Wanapu, 2013):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

where  $A$  and  $B$  are keyword sets regarding two scientific papers and  $J(A, B)$  denotes the Jaccard similarity between them. In other words, an edge was drawn between two nodes  $v_A$  and  $v_B$  if  $J(A, B) > 0$ .

Once the network of papers was drawn, community detection was used to examine the structure of the network. At this step, the *Walktrap* approach of Pons and Latapy (2005) was used because it captures the community structure properly with efficient computing and maintains an agglomerative algorithm to determine the optimal number of communities for the purpose of interpretability (Pons & Latapy, 2005). The *Walktrap* method allows random walk across the network to calculate the distance between  $v_A$  and  $v_B$  based on the probability that a random walker moves from  $v_A$  and  $v_B$  with a fixed number of steps  $t$ . Nodes can be merged into communities with hierarchical clustering based on the distances between the nodes. The distances between nodes should be considerably shorter within communities than between communities.

If the results of community detection and LDA topic modeling are in line with each other, the reliability of our findings and conclusions are further strengthened.

## 4. Results

### 4.1. Answering the research questions

Due to the original research questions, the goal was to find the extracted topics of SCAI-related studies along with their networks, hierarchical relationships, and timeline with future perspectives. In this section, the aforementioned research questions will be answered using the results obtained via LDA topic modeling and community detection. To find the leading topics as signals of research trends, text pre-processing was applied, including tokenization, lowercase transformation, removal of special characters and stop-words, and lemmatization, before the LDA model was used. The hyperparameter selection was optimized  $\alpha$  with the built-in optimizer of R's "textmineR" package. Different parameter levels related to the  $\beta$  parameter were selected for model candidates, and qualitative investigation was applied to fine-tune the model (Maier et al., 2018).

Three different approaches to the selection of an optimal topic number were used (Arun et al., 2010; Cao et al., 2009; Griffiths & Steyvers, 2004), as described in Section 3.2. Fig. 2 shows the result of the simulation.

The simulation was conducted; for each iteration, the number of topics was changed, and the three mentioned metrics were recalculated. The upper part of Fig. 2 shows the objective functions that need to be minimized, while the lower diagram presents the objective functions that need to be maximized. Based on the results, the selected topic number should be between 15 and 18, since the



**Table 1**  
Summary of the extracted topics.

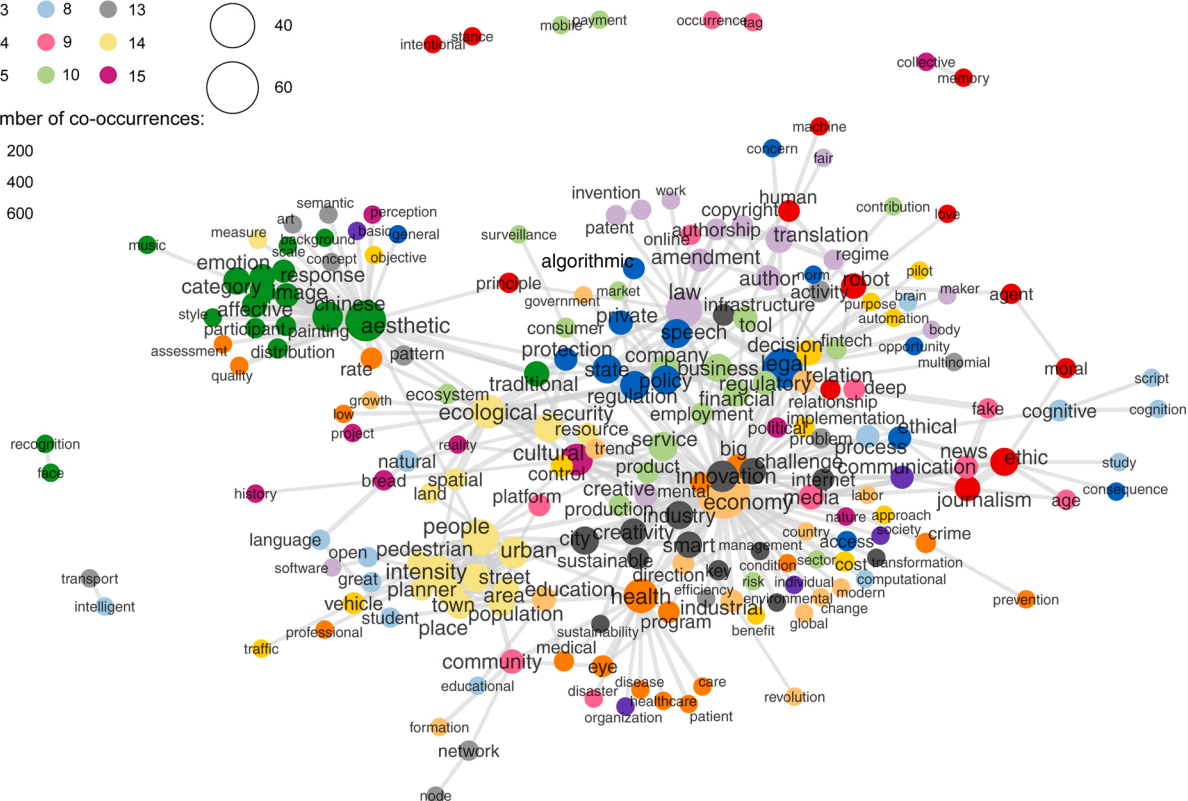
Topic id	Label	Number of papers
1	Governing AI and policy making	47
2	Human-robot communication	58
3	AI-driven art	32
4	Social communication, complex systems	43
5	Healthcare	35
6	Smart city, sustainability	47
7	Automation, vehicles and networking technology	28
8	Education, natural language	53
9	Social media, fake news/deep fake	48
10	FinTech, employability, business	28
11	Patent, copyright, trade secret	27
12	Industry 4.0, information technology	55
13	Social economy, interactions	37
14	Urbanisation, work-force, security	23
15	Virtual reality, cyberconstructions	46

Legend:

Colors by topic id:



Number of co-occurrences:



**Fig. 3.** Network of terms based on co-occurrence within same sentence (considering the entire corpus).

metrics given by [Cao et al. \(2009\)](#) and [Arun et al. \(2010\)](#) have their minimum within this interval and the function based on [Griffiths and Steyvers \(2004\)](#) flattens after  $T^* = 18$ .

To interpret the results, the terms with the highest probabilities for each topic were reviewed; sample documents representing a high proportion of specific topic were also read through, as suggested by [Quinn, Monroe, Colaresi, Crespin, and Radev \(2010\)](#), [Jacobi, Van Atteveldt, and Welbers \(2016\)](#) and [Maier et al. \(2018\)](#).

After qualitative investigation (Maier et al., 2018) of different topic numbers from the highlighted interval, fifteen extracted topics

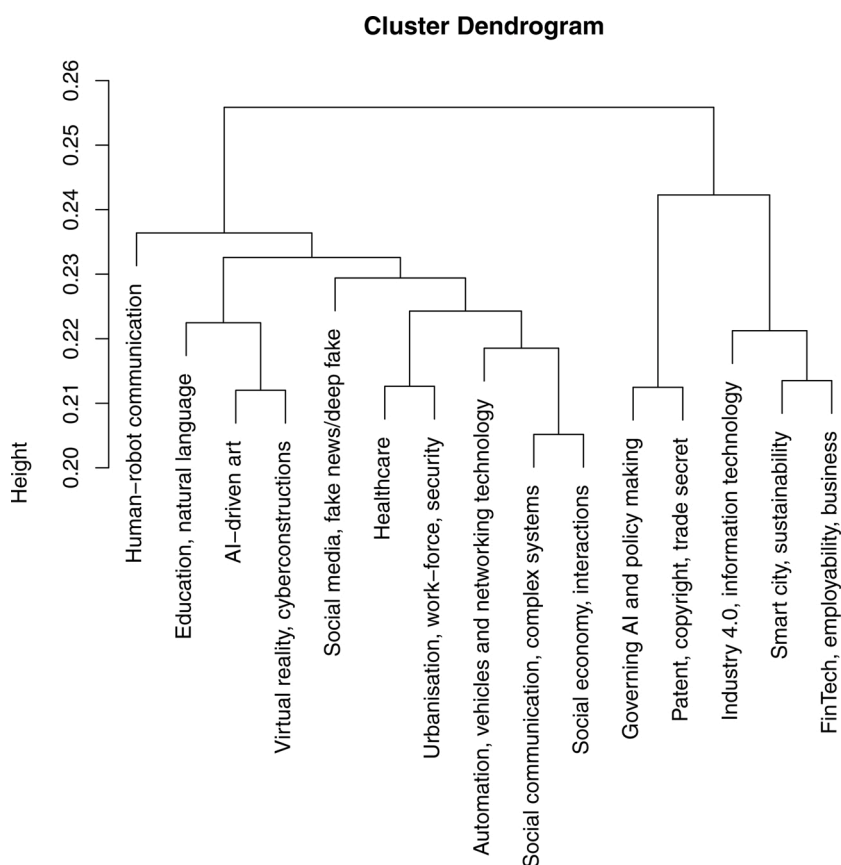


Fig. 4. Structure of the extracted topics given by Hellinger distance.

were clearly identifiable as having optimal quantifying. Following the aforementioned research methods, each single topic was limited to a minimum of one and a maximum of three sub-themes. When the sub-themes of each extracted topic were consolidated this way, the number of leading topics became fifteen.

The fifteen extracted topics of SCAI are the answers to RQ1 (see Table 1). Well-established fields like regulation, business, education, communication, and media were among the extracted topics, along with innovative fields, from FinTech to cyberconstructions to deepfakes. Although top-down governmental and business decision-making is also in the spotlight, as was expected, topics related to bottom-up stakeholders, such as communities, civic engagement, or NGOs, are absent from the key list. These results show the priority of top-down innovative efforts in SCAI research.

As Table 1 summarizes, broad and complex fields are confirmed on the extracted list, covering almost all social and cultural areas. According to the results, the most frequent topic is human-robot communication. This result highlights the intensive research in the changing relationship between mankind and technology, and also confirms the need for snapshot research about SCAI.

One part of the extracted topics is congruent, such as governing with policymaking or patent with copyright and trade secrets. Further ones are diverse, for example, urbanization with workforce and security, or FinTech with employability and business. Therefore, well-defined and even fuzzy categories are available in parallel as different approaches of social structures and cultural meanings. To understand these fifteen shadows of SCAI with the congruent and diverse results, the next step is to identify the interconnected topic areas in detail.

Further investigation revealed a dense topic network (Fig. 3). In Fig. 3, nodes represent the extracted terms and edges denote the frequency of their co-occurrence. The nodes are colored based on their topic membership according to the LDA method. The presented network defines the center, peripheries, and sub-graphs appropriately and points out the inherent interdependence and system-wide relationships of each extracted topic.

Fig. 3 shows the influential nodes with expected future expansion and less-connected themes and answers RQ2; the co-occurrence of the same sentence has revealed a densely topic network, which reinforces the aforementioned close association between each output shown in Fig. 3.

Interpreting Fig. 3, several directly connected centers are revealed, such as economy and innovation, law and regulation, security and control, people and urban area, or aesthetics and emotions. Social sciences are mostly represented this way and humanities are underrepresented. This result was expected in line with the validation tests when cultural aspects were less highlighted in the studied databases compared to the social factors (see Section 3.2). However, it is a noticeable finding that most of the subject areas of SCAI



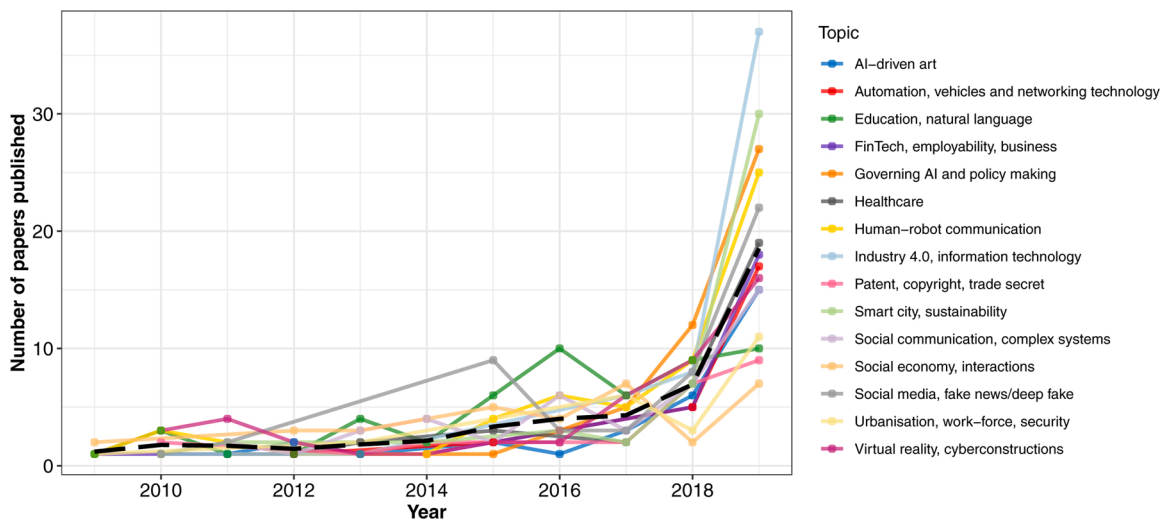


Fig. 5. Published papers in each topic by year.

research are closely interlinked in one dense network. In contrast, aesthetics-related terms densify into another subgraph with own trending topics from emotions to affected participants. Presumably, there is a double reason for these results. First, social and cultural topics are mostly multi-related along with AI technology, confirming the importance of introducing the term SCAI. Second, a few culture-related studies need more time to be developed and integrated into the multiconnected SCAI concepts with perception, intuition, or well-being.

Nodes with a high degree present intensity, creativity, or challenge for the majority of SCAI research as changing techno-AI developments. However, the central node is “economy”. This is one of the most important results obtained, pointing out that the economy becomes the main driver for SCAI research beyond technology. To sum the RQ2 with future perspectives, techno-economic innovations will presumably drive and shape the landscape of subject areas with even more interconnected fields and the cultural issues will be investigated more deeply.

The connected parts in the periphery represent diverse fields. Most of the topics belong mainly to human-related themes, such as music, moral issues, history, or love. However, further peripheries also highlight techno-AI topics, namely machines, networks, agents, or cognition. All these peripheral topics are probably about fine-tuning the human-machine relationship in the future. For example, Lovotics is presented in an article for a relation between intimate human love and AI.

The mostly converging topics are noticeable as RQ2.1. presenting strong signals of research trends. The strongest connections are evident between “economy and innovation” or “law” and “regulation”. These mostly represent social science, but also partly computer science and engineering. Less central but converging topics, such as city characteristics with population, the health industry with eye care, and journalism with ethics extend these with several subject areas and disciplines. It is unexpected that “culture” is strongly integrated and interlinked with central nodes, although humanities are underrepresented in the entire database. This result can be explained by the fact that culture is multi-related to several diverse areas which gives it a serious potential for further research.

RQ2.2 asked which topics are not directly connected presenting weak signals of emerging research trends. The answers are divided into two groups. The first group represents applications in practice, such as facial recognition, mobile payments, or intelligent transport systems. The second group invites different disciplines to discuss complex developments, from intentional stance to collective memory. These detached peripheries are not without associations but are only tightly connected to the fifteen extracted topics. These isolated subgraphs have a strong characteristic in techno-AI, therefore their direct links to the leading topics are expected, in the near future, to support inter-disciplines.

To find the answer for RQ3, a hierarchy of the extracted topics was revealed using Hellinger distance (see Fig. 4):

$$H(p, q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{N_v} (\sqrt{p_i} - \sqrt{q_i})^2} \quad (3)$$

$N_v$  represents the number of terms in vocabulary  $V$ ,  $p$  and  $q$  denote probability distributions regarding two topics.

The dendrogram confirms the interplay between techno-AI and SCAI with multiple branching. The illustration presents the topic hierarchy as (1) diverse SCAI-related themes with techno-economy and human-robot communication; (2) techno-AI with governmental, industrial, business, economic, urban, regulative, and sustainable research. This result confirms the introductory theoretical proposal to distinguish SCAI, and techno-AI research (see Section 1). It is also visible how computer science and engineering are converging to social science. With all these, it is techno-AI research that will explore SCAI issues more deeply with a high probability in the short run. This will be confirmed in the next section by the Fig. 5 too.

AI with automation, communication, and interactivity features can support certain basic societal functions (Sundnes, 2014),

**Table 2**  
Policy-related content.

Topic	Count	Policy	Level
Governing AI and policy making	11	Policy research, policy process, policy framework	State/national
Smart city, sustainability	10	Policy-making improvement, concepts and predictions for policy making	Local or regional
FinTech, employability, business	9	Policy reports, policy planning, policy design	State/national and union
Healthcare	7	Evidences for policy making and policy for social investment programs	Governmental and NGO
Patent, copyright, trade secret	7	Policy research, policy recommendations	State/national
Automation, vehicles and networking technology	6	Public policy recommendations, policy analysis	Governmental
Urbanisation, work-force, security	5	Policy design, urban policy	Local and regional
Industry 4.0, information technology	4	Public policy, industrial policy	State/national
Human-robot communication	3	Evidence-based policy making, policy frameworks and shared policies	Institutional and ownership-based
Social communication, complex systems	3	Public policy, policy design, evidence-based policy making	Business
Social media, fake news/deep fake	2	Corporate policy, public policy, guided policy	Corporate and user
AI-driven art	0	N/A	N/A
Virtual reality, cyberconstructions	0	N/A	N/A
Education, natural language	0	N/A	N/A
Social economy, interactions	0	N/A	N/A

including education, healthcare, urban utilities, engineering, transportation, security, and social regulation. However, further societal functions, such as logistics, food-related topics, shelter, and clothes are underrepresented or invisible in the scope. Considering these functions, the theoretical yin-yang model above (Seel, 2012) is partly confirmed as the dynamic connection between technology and its social environment.

The cultural functions (Ting-Toomey & Dorjee, 2018) are also represented on different levels. The most significant of these functions are art and design with research and projects as well. Media, identity constructions, and communities are partly available via social communication and natural languages. This result is consistent with the underrepresented cultural studies in our database. Natural language processing is assumed to be as fundamental as human-robot communication Deng and Liu (2018) but it is still subordinated to larger topics. This trend can change rapidly with upcoming developments of techno-AI. Research into human-robot communication can facilitate this direction.

Investigating SCAI in general, the least diverse fields are authorities, policy- and decision-making, with a converging cluster consisting of governing AI, trade secrets, and copyright. Regulation is all around. Beyond these, topics represent diversity and inter or multidisciplinary. Mostly socio-economic aspects influence the development of AI and vice versa. This result is particularly interesting if the keyword “culture” is paired with “society” to build the data collection. Therefore, the sociocultural dimension should be extended to business and finance, assuming a multidisciplinary research practice. From this perspective Seel’s tao model can also be extended.

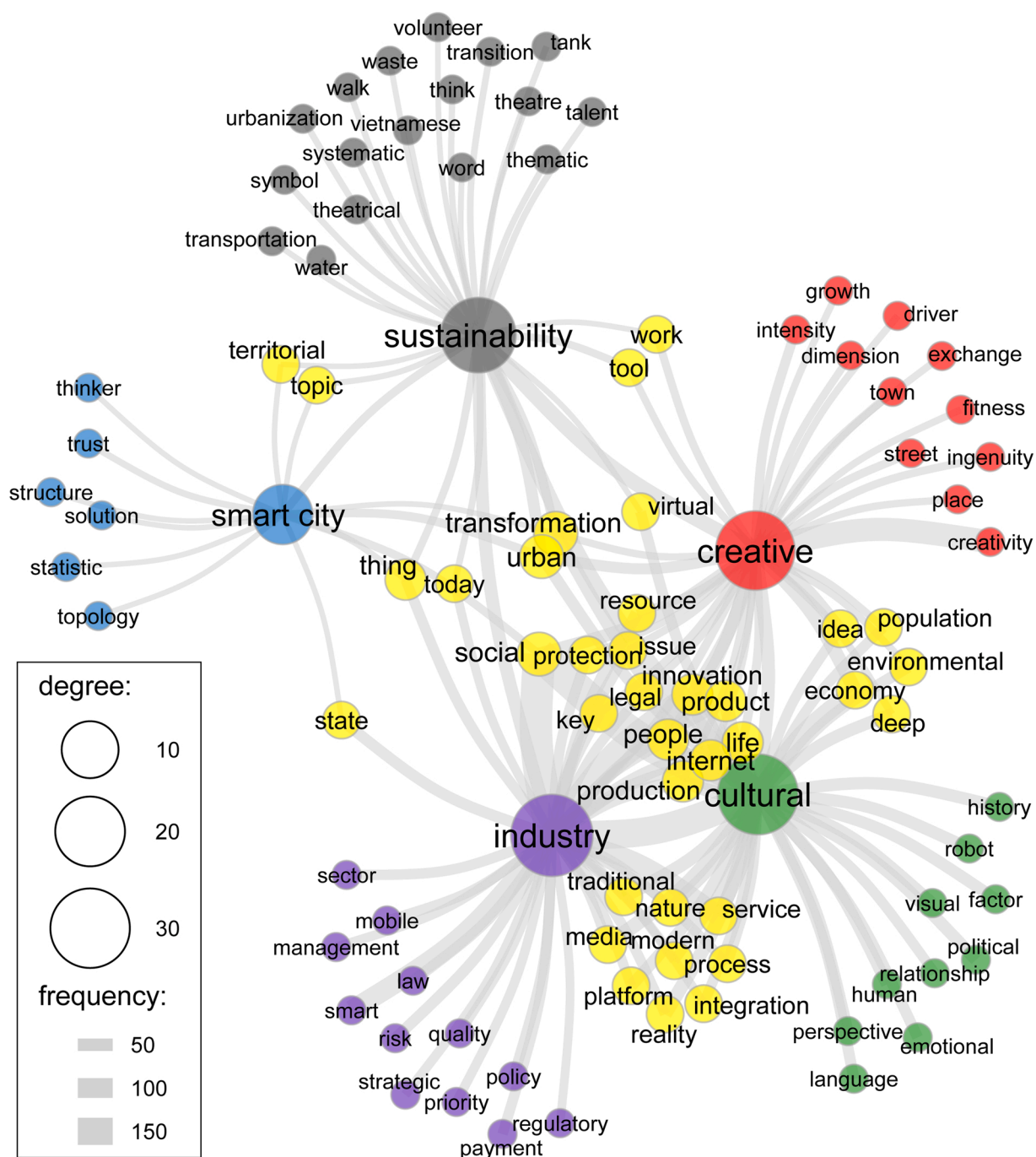
As for RQ4, the timeline of the emerging research topic trends expands the level of analysis. Looking back to the last ten years, significant growth begun in 2015 with social or fake media and continuing in 2016 with education or natural language (Fig. 5).

As we expected due to the theoretical considerations, a change point year of 2018, for SCAI research was discovered. Since then, SCAI research has been increased dramatically, mostly in Industry 4.0, smart cities, governing AI, and policymaking, but essentially in most topics, reflecting the rapidly emerging techno-AI challenges. With the leading topics, engineering, computer science, and policy research support social science and humanities with future perspectives. Although slower trend changes are detected in research of social economy, education, or regulation, these have boomed several times over the past decade. All these trends of publications indicate (dashed line on Fig. 5) further intensive growth of the extracted topics, mostly for social science and with more focus on engineering.

Answering RQ5 we also investigated SCAI research trends for policies. The drive of their trends are social-cultural forces that discover the changing technology (Dragt, 2018) at indifferent policy levels. The vast majority of extracted topics present “policy” mentions in different forms but only ten percent of all the analyzed abstracts contain policy-related subject areas (Table 2).

This is a niche for SCAI research and also for policy research. Only a few mentions of “policy” are highlighted in several abstracts mostly with evidence-based policymaking or policy design. From policy types, public policy is the only one highlighted confirming the above revealed key role of governing AI. Industrial or corporate policies are hardly mentioned. These results for SCAI and policy research should be considered as to how they can be proactive in supporting policymaking.

The topics with the highest frequency of policy-related terms are mostly human-robot communication, automation, Industry 4.0, urbanization, cybersecurity, and social or fake media. The key subject areas of these are trustworthy technology and AI ethics, sustainability, changing labor market, standards for social values, democracy, and regulation. Beyond these emerging fields, the further policy mentions and their research topics are sporadic, mentioning different types, fields, and contexts of policies. Related cultural issues are hardly represented. Considering these results, SCAI research has only weak and partly defined links to policymaking. Although the existing abstracts present proactive ideas, policy recommendations are almost completely missing and mitigation plans are unavailable for social or cultural issues. Based on these results, expansion of policy-related research is expected in SCAI, and research funding is proposed to investigate the social-cultural needs and values along with “digital trust” with secure, ethical, and

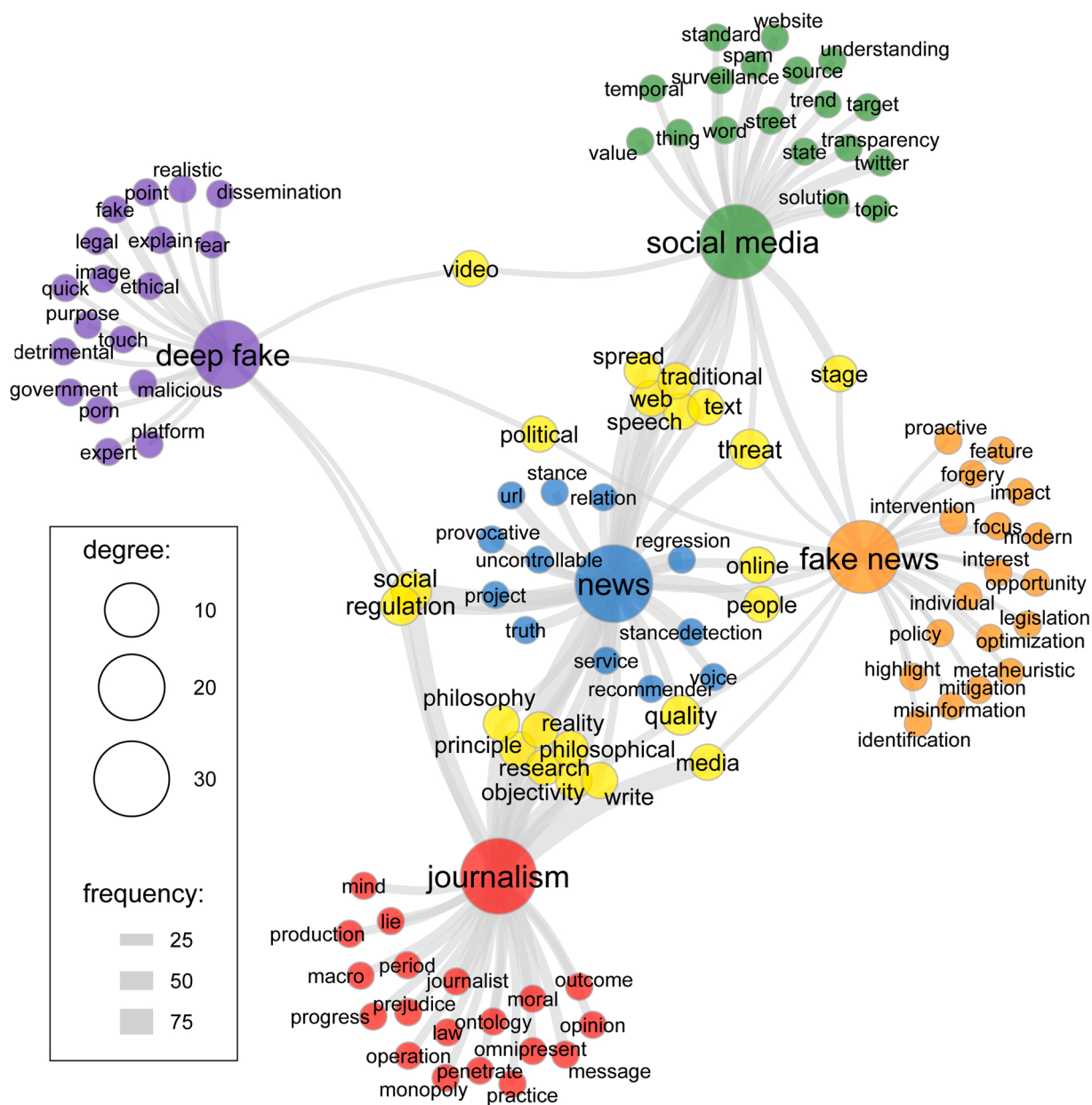


**Fig. 6.** Co-occurrence network of terms related to “smart cities” within same sentence. Yellow nodes represent the intersection between the analyzed terms.

reliable technology (Roberts et al., 2021), (Vesnic-Alujevic, Nascimento, & Pólvor, 2020), (Robinson, 2020), (Diallo et al., 2021), (Schürer, Stangl, Müller, & Hubatschke, 2017). Techno-AI developments invite SCAI research to support policy-making this way.

#### 4.2. Associated themes and weak signals of the future

Along with the analysis of the fifteen extracted topics, certain contextually-associated keywords were also revealed beyond the original research questions. Therefore, a co-occurrence network was provided by Fig. 3 to select the keywords of interest with their association network. The first branch of these fields contains creative industry, cultural industry, and creative-cultural industry, with a



**Fig. 7.** Co-occurrence network of terms related to “journalism” and “social media” within same sentence. Yellow nodes represent the intersection between the analyzed terms.

strong connection to sustainability and smart cities (Fig. 6).

Therefore, research on culture and creativity converge on the subject of smart technology in urban contexts (Hatuka, Rosen-Zvi, Birnhack, Toch, & Zur, 2018) to create valuable and liveable places. Creativity drives intense social-business growth, while cities provide structural space for growth through strategic management and policymaking. Human-oriented innovation and production are the drives of cultural-creative industries, while digital transformation supports urbanization through IoT and internet services. SCAI research in these areas tends to emphasize the importance of talents and volunteers as well as human emotions and relationships at the level of the individual. This field presents all studied disciplines, such as computer science, engineering, social science, and humanities with highlighted cultural studies, thus the most complex joint techno-AI with SCAI.

The cultural-creative industry and smart cities with sustainability issues are converging on each other without any central node in Fig. 6. If this result is compared to the outputs of the dendrogram (Fig. 4), language and visual communication are highlighted here only in the form of the basic cultural functions mentioned above (Ting-Toomey & Dorjee, 2018). Besides, techno-AI is shown to be well-connected to culture through robotics. Consequently, the essential human-robot communication is highlighted in this

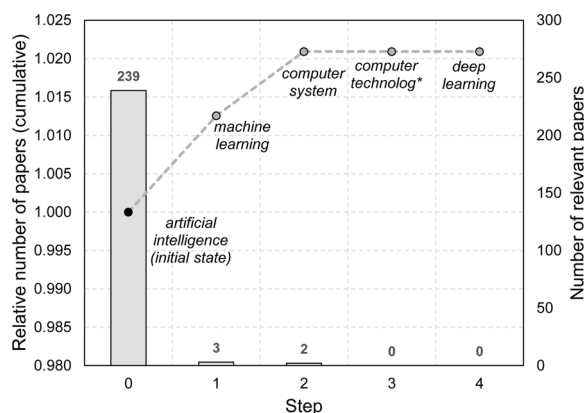


Fig. 8. Cumulative ratio of relevant papers according to the use of additional keywords. Black dot represents the initial state.

co-occurrence network as well.

The term “theatre” in the subnetwork of sustainability was an unexpected node in the co-occurrence network. In the original database, “theatre” is also associated with robotics in the context of robot-written music or plays, and with robots implemented in performances or for social sustainability (Eizenberg & Jabareen, 2017). These kinds of specific cultural approaches and applications are expected to be expanded with many other cases and best practices in the next decade in different cultural industries.

Second, a closely-related and still-branching area is journalism and social media, along with fake news and deepfakes (Fig. 7).

These intersecting areas describe the contemporary media landscape, tackling issues like the news industry vs. misleading information or AI-driven face swaps in videos (Shae & Tsai, 2019; Stover, 2018). It is noticeable that this theme network provides the longest connection across the fifteen extracted topics around the SCAI themes, penetrating several topics.

The absolute center of the co-occurrence topic network is the node of “news,” which is mostly connected to “journalism” and to fundamental questions of reality and quality. The news distributor “social media” is also strongly connected with critical aspects of transparency and surveillance. Both “deepfakes” and “fake news” are linked to this because these terms suggest misinformation that can quickly reach millions of people or methods of social regulation.

The key contents of this branching area are political campaigns and revenge porn. In these contexts, it is crucial to identify misleading information and to find proactive solutions to avoid these problems or mitigate their impact. It also confirms the above mentioned problem with missing mitigation plans in policy topics. Organizational, business and public policies will play a critical role in this. These results confirm the pervasive role of AI-powered media in society and culture.

The changing balance of media becomes visible in Fig. 7. Journalism and social media created a massive flow of information originating from editorial boards or users. In the meantime, techno-AI started to convey automated fake news and deepfakes to large masses in the short term or in real-time. Accordingly, verifiable news and deepfakes refer to an imbalance in the media industry that is worth exploring in more depth.

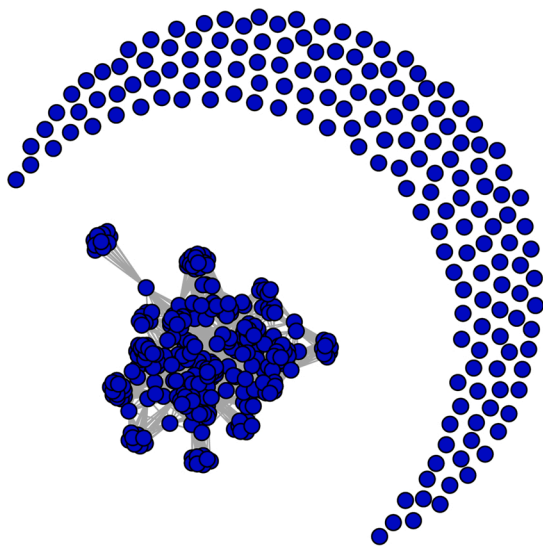
In this section we also focus on the detection and discussion of weak yet relevant signals. As Lee and Park (2018) and Burmaoglu, Sartenaer, Porter, and Li (2019) suggest, this type of quantitative analysis can also be used to identify potential drivers of changes via emerging topics or discussions. Our quantitative approach supports the detection of these weak signals since co-occurrence networks are not only useful to highlight edges with the greater weights but also make it possible to observe the weak but noticeable patterns within the corpus. In our case, edges between the central nodes with lower or higher degree are associated to this concept, see the yellow data visualization on Figs. 6 and 7.

Starting with the low degree nodes, “state”, “territorial topic” (Fig. 6) and “social regulation” (Fig. 7) confirm again the growing role of regulation and policy making. Only the democratic values determine this direction in political context according to the investigated articles. The emerging information tools and even more users online force this direction. The sensor-audio-video based data analysis and the upcoming AI development stages are also change drivers. The topic with the lowest degree is “video” between social media and deepfakes, presenting vlogs, video and brain games or hyper-realistic contents. This weak signal predicts that amount of automated video contents continue to grow intensively and fundamentally affects the interpretation of reality with extending fake media.

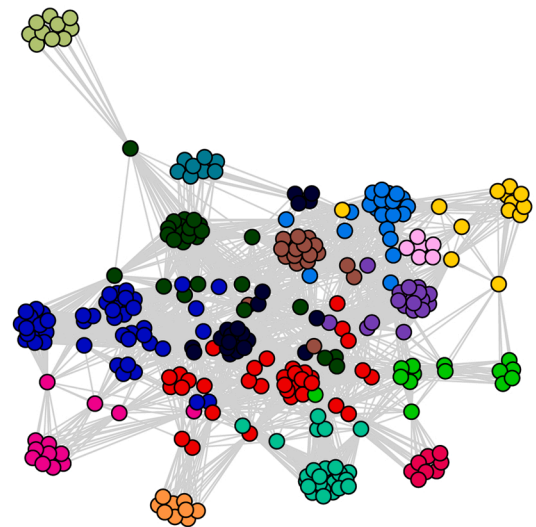
Interpreting the edges with high degree, both co-occurrence networks summarize complex issues to emerge. The most noticeable is that concept of “reality” (Figs. 6 and 7) is re-evaluated in context of ethical-legal safety and social values. This approach assumes that complex scenarios will be needed to build trust in AI-generated reality. It is noticeable that the changes in media industries imply more fears while smart city concepts represent mostly competitive ideas. In this case, one optional scenario is an increasingly regulative society against misinformation and another one scenario is a “data and idea” based ambitious economy.

These drivers of changes and the relevant signals are useful for future studies, technology foresight and policy making in predictive analysis and scenario planning.

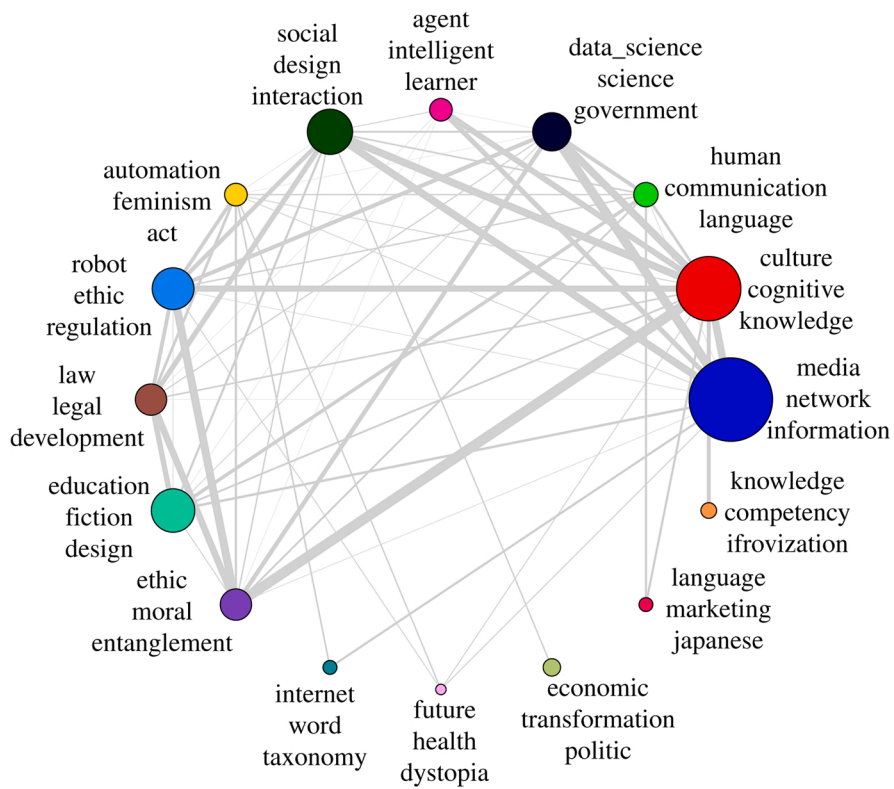




(a) Entire network



(b) Non-isolated nodes after community detection



(c) Aggregated graph

Fig. 9. Results of the community detection.



**Table 3**  
Summary of RQs, results and implications.

Number	RQ	Result and implication	Table/ Figure
RQ1	What are the leading topics of SCAI in academic research?	RESULT: Fifteen extracted research topics of SCAI were defined mostly from social science and driven by techno-economy. The importance to introduce SCAI concept was confirmed. IMPLICATION: Well-defined research areas have become available, presenting the broad impact of AI technology. Contextual integrity and extension of humanities are proposed to interpret social values of the future.	Table 1
RQ2	Which SCAI research topics are the most influential and which are on the periphery with less focus?	RESULT: The most interconnected center is economy and innovation extended directly with regulation, urban topics, culture and control and remotely with aesthetics and emotions. IMPLICATION: Grants and funding should support these central topics to understand the key social-cultural issues in AI technology and art projects can encourage the detachment centers to be integrated to interpret the changes.	Fig. 3
RQ2.1	What are the core converging SCAI research topics?	RESULT: Regulation, urbanization, health care and media present strongly associated emerging research trends. IMPLICATION: The role of the social sciences is strengthened, linked to computer science, engineering and medicine. Their inter- and multidisciplinary research should be supported for effective and valuable developments.	Fig. 3
RQ2.2	Which SCAI research topics are not directly interconnected and are on the periphery?	RESULT: Applications, such as facial recognition or mobile payments, and also complex developments with collective memories are remote topics. IMPLICATION: Probably, the barriers of the techno-AI adaptation resulted in this finding. Its specific fields should be more discovered for policy or ethical research.	Fig. 3
RQ3	What is the hierarchical relationship between the leading SCAI topics?	RESULT: The human-robot communication is the cross-topic for SCAI. Further diverse fields are described by governing AI, regulation, Industry 4.0, smart urbanization and business-financial developments. IMPLICATION: The focus is on the human-machine ecosystem extending broadly in society and culture. The diverse research fields will probably join this movement. Investments should focus on ecosystem-based innovation.	Fig. 4
RQ4	What have been the trending SCAI topics over the past decade with expected future expansion?	RESULT: The trends are changing but year 2018 is the tipping point for SCAI research to grow dramatically, mostly with Industry 4.0, smart cities and governing or policy issues. IMPLICATION: The found tipping point assumes rapid growth also in short and long term for SCAI. Engineering and computer science facilitate these movements along with techno-AI. Social sciences, humanities and policy making should be more proactive keep up the pace.	Fig. 5
RQ5	What are the SCAI research trends for policies?	RESULT: Majority of the extracted topics highlight the challenges of policy making, mostly with issues in trust, sustainability, ethics, regulation and democracy. IMPLICATION: Policy recommendations should be more represented as well as the industrial or business policy frameworks. This will be a challenge for policy research and a pressure by policy making.	Table 2

## 5. Validation of keyword search and model results

### 5.1. Validation of the keyword search

To validate the keyword search, we applied the approach described by Section 5.1. The analysis was conducted considering the relevant papers collected from Scopus because it is the largest set from all the data sources we used. The time span was the same as described in the PRISMA statement. The list of additional keywords is defined as: “*machine learning*”, “*computer system*”, “*computer technology*”, “*deep learning*”. It is important to mention that an asterisk is applied in the case of “*computer technology*” to ensure papers containing the word “*computer technologies*” are also collected. Fig. 8 shows the result of the sensitivity analysis.

In Fig. 8, the left axis shows the cumulative ratio of relevant papers found and the right axis presents the absolute number of papers by each step. As the results show, although the addition of keywords “*machine learning*” and “*computer system*” would slightly increase the number of relevant papers, extending the initial state with the defined list would only provide an additional 2% of relevant papers. Therefore, the result from the initial search rule is not sensitive to additional keywords and using “*artificial intelligence*” as a keyword in the search leads to valid results.

### 5.2. Validation of the topic modeling results

Beyond the investigated abstracts, the keywords of the articles were aggregated into a graph. To validate the results given by LDA

topic modeling, a keyword-based network community detection method was used, as described in Section 3.5. Fig. 9a shows the network of papers, including isolated nodes. It represents the collected papers, while edges point out the Jaccard similarity between them. Fig. 9b presents the results of community detection after removing isolated nodes. Finally, the better structure of the results with a community graph is provided (Fig. 9c), where each node represents a community, and edges denote the number of connections among them. The size of each node reflects the number of papers included by the given community, and the labels describe the most frequent keywords.

The first important observation is that the number of extracted topics are in line with the optimal number of LDA themes. While the simulation suggested that 15–18 topics should be specified, community detection clustered the papers (based on keywords) into 16 groups. The entire network (Fig. 9a) shows significant diversity, as in the case of the abstracts. However, a strongly connected hub has also become available when sixteen topics are extracted. This output presents a massive association with the abstract themes given by LDA.

As Fig. 9b shows, topic communities are clearly detectable. Fig. 9 confirms that the key nodes with the highest significance are “media,” “network,” and “information,” as was discussed several times above. It seems that this finding also confirms that “media” is the most cross-cutting and reflexive field. Besides, data science and data governance are mostly associated with social networking and information flow in the contexts of policy types, regulation, and ethics. The low focus on policymaking or policy recommendations is confirmed with this validation as well. This niche should be discovered in near future. Research in culture is also more represented in the case of keywords. This focus can be improved by the validated categories of knowledge, cognitive functions, and moral issues. Keyword mining of community networks additionally resulted in the visibility of further topics, namely marketing, dystopia, feminism, and singularity in the socio-cultural context, as well as the importance of blockchain to techno-AI. Only the term “innovation” forms a bridge between the sixteen nodes. All substantive results, however, reinforce the analytical results from the abstracts.

## 6. Conclusion and discussion

This research revealed the fifteen leading topics of SCAI research using complex association networks, hierarchy, and timeline, which were obtained via LDA topic modeling and community detection, beyond the approach of techno-AI. The revealed association networks also confirmed the key results. Consequently, the study achieved its goal to introduce a systematic review of SCAI research trends from seven academic databases with a decade-long time span presenting key topics and implications for academia and policymaking. This comprehensive summary also extends and interprets the theoretical considerations of Seel’s Jin-Jang model and its limits (Seel, 2012).

According to the results, innovation economy and human-robot communication mostly connect the significant topics for SCAI research. However, several research fields are interconnected strongly. Therefore, inter and multi-disciplines discover techno-AI, mostly by social science along with computer science and partly humanities, engineering, and medicine. Topics of humanities are sporadic in many cases but a weakly connected module of cultural studies is represented by art and emotions. The most complex and associated field is “smart city” with creative-cultural industries, while media play a pervasive role in SCAI research. A more detailed investigation is needed in these areas to understand the complexity of the constantly evolving human-machine ecosystem. The results and their implications are summarized as follows with the original RQs and the list of illustrations (Table 3).

Considering the implications and results, techno-AI with ever wider applications and fifteen shadows of SCAI research predict the upcoming machine-human ecosystem and the related research responsibility. Therefore, SCAI research fields project mostly the high degree of complexity from regulation to economy or from smart city to media covering a wide spectrum of socio-cultural functions and dilemmas. The change point of this future perspective was found in 2018 confirming the proper timing of our snapshot research. The fine-tuning of machine-human relations has started, thus human needs and values have started to investigate in an AI context. Non-biased environments (Noriega, 2020) risks of superintelligence (Brundage, 2015) or fake media call for research to study the issues in trust, sustainability, regulation, and democracy. These issues are expected to remain key focus areas in the long run. This result is consistent with the found niche in policy research or weak connection to policymaking almost without frameworks or recommendations.

The introduced concept of SCAI and the revealed research trends will support academic and policy decisions in case of investments and funding in several ways, such as (1) finding socially-culturally acceptable adaptation process (2) defining issues, biases, and mitigating proactively (3) revealing needs to change (4) presenting insight for policy making (5) highlighting the values and advantages of AI technology. Moreover, the summarized weak signals support understanding not just the simple predictions from extending video-based reality to competitive idea-economy, but also recognizing the deep complexity of future AI with ontological or regulative issues. In line with these, SCAI can also reveal the bottom-up research alternatives, namely, how policy can support an empowered individual or community to adapt to the upcoming techno-AI in society and culture. The key challenge is the effective harmonized adaption in the context of SCAI.

## 7. Limitations

The present research has three basic limitations. First, this study did not aim to analyze or visualize all fields as it only focused on the future direction of key emerging fields. Second, the research is not repeatable as the algorithms and indexing methods behind the academic databases are changing. Therefore, the ambition of this study was only a snapshot to understand the research trends of SCAI. Third, Google Scholar was excluded from the databases studied. The reason for this was that only the comparable academic databases were applied for the highest-ranked publications. Based on the above mentioned points the further improvement of research design

**Table 4**

Top five terms in each topic based on the calculated  $\varphi$  values ( $t_i$  denotes the id of each topic given by Table 1).

$t_1$		$t_2$		$t_3$	
Term	$\varphi$	Term	$\varphi$	Term	$\varphi$
legal	0.0059	robot	0.0141	emotional	0.0033
law	0.0027	social	0.0045	aesthetic	0.0024
regulation	0.0023	human	0.0034	emotion	0.0022
ethical	0.0021	ethic	0.0030	cultural	0.0020
policy	0.0021	machine	0.0025	music	0.0019
$t_4$		$t_5$		$t_6$	
Term	$\varphi$	Term	$\varphi$	Term	$\varphi$
social	0.0068	health	0.0058	smart	0.0039
dynamic	0.0016	medical	0.0022	innovation	0.0029
communication	0.0015	patient	0.0021	cultural	0.0026
society	0.0015	healthcare	0.0016	industry	0.0026
organization	0.0015	social	0.0016	city	0.0025
$t_7$		$t_8$		$t_9$	
Term	$\varphi$	Term	$\varphi$	Term	$\varphi$
decision	0.0034	cognitive	0.0041	media	0.0074
autonomous	0.0023	language	0.0040	social	0.0064
automation	0.0021	student	0.0037	social_media	0.0045
vehicle	0.0019	learner	0.0018	news	0.0027
control	0.0015	natural	0.0018	people	0.0021
$t_{10}$		$t_{11}$		$t_{12}$	
Term	$\varphi$	Term	$\varphi$	Term	$\varphi$
financial	0.0028	law	0.0038	economy	0.0064
service	0.0028	author	0.0019	social	0.0045
business	0.0024	property	0.0018	education	0.0025
product	0.0018	copyright	0.0015	industrial	0.0024
consumer	0.0017	work	0.0015	change	0.0016
$t_{13}$		$t_{14}$		$t_{15}$	
Term	$\varphi$	Term	$\varphi$	Term	$\varphi$
social	0.0031	people	0.0026	cultural	0.0028
network	0.0030	urban	0.0025	history	0.0015
high	0.0013	place	0.0019	century	0.0014
similarity	0.0012	ecological	0.0016	political	0.0012
state	0.0010	intensity	0.0016	virtual	0.0012

and applied methodology is limited by the shortcomings of data sources and the snapshot nature of the data. Despite all these limits, this study was able to find the trending research topics of SCAI and support implications to future perspectives.

#### Declaration of interests

None.

#### Acknowledgements

Sincere gratitude to the Fulbright Research Grant for the opportunity to work on this project. This paper was also supported by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the TKP2020-NKA-10 project financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme from the National Research, Development and Innovation Fund of Hungary.

#### Appendix A. Most representative terms per topic

Table 4

## References

- [Arun et al., 2010] Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 391–402).
- [Bloom, 2020] Bloom, P. (2020). Legal reboot: From human control to transhuman possibilities. In *Identity, institutions and governance in an AI world* (pp. 211–246).
- [Brand et al., 2020] Brand, M., Jirtoka, M., Webb, H., & Luff, P. (2020). Fair navigation planning: A resource for characterizing and designing fairness in mobile robots. *Artificial Intelligence*, 103259.
- [Brundage, 2015] Brundage, M. (2015). Taking superintelligence seriously: Superintelligence: Paths, dangers, strategies by Nick Bostrom (Oxford University Press, 2014). *Futures*, 72, 32–35.
- [Burmaoglu et al., 2019] Burmaoglu, S., Sartenauer, O., Porter, A., & Li, M. (2019). Analysing the theoretical roots of technology emergence: An evolutionary perspective. *Scientometrics*, 119, 97–118.
- [Cao et al., 2009] Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72, 1775–1781.
- [Chen et al., 2017] Chen, P., Zhang, N. L., Liu, T., Poon, L. K., Chen, Z., & Khawar, F. (2017). Latent tree models for hierarchical topic detection. *Artificial Intelligence*, 250, 105–124.
- [Coeckelbergh, 2020] Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- [Deng and Liu, 2018] Deng, L., & Liu, Y. (2018). *Deep learning in natural language processing*. Springer.
- [Di Vaio et al., 2020] Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, 121, 283–314.
- [Diallo et al., 2021] Diallo, S. Y., Shults, F. L., & Wildman, W. J. (2021). Minding morality: Ethical artificial societies for public policy modeling. *AI & Society*, 36, 49–57.
- [Dragt, 2018] Dragt, E. (2018). *How to Research Trends: Move beyond trend watching to kickstart innovation*. BIS Publishers.
- [Dwivedi et al., 2019] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., et al. (2019). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994.
- [Eizenberg and Jabareen, 2017] Eizenberg, E., & Jabareen, Y. (2017). Social sustainability: A new conceptual framework. *Sustainability*, 9, 68.
- [Emmert-Streib et al., 2019] Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2019). Utilizing social media data for psychoanalysis to study human personality. *Frontiers in Psychology*, 10.
- [Feher and Zelenkauskaitė, 2020] Feher, K., & Zelenkauskaitė, A. (2020). *AI in society and culture: Decision making and values*. arXiv preprint arXiv:2005.02777.
- [Felzmann et al., 2019] Felzmann, H., Villarronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6, 2053951719860542.
- [Galanos, 2018] Galanos, V. (2018). Artificial intelligence does not exist: Lessons from shared cognition and the opposition to the nature/nurture divide. In *IFIP international conference on human choice and computers* (pp. 359–373).
- [Griffiths and Steyvers, 2004] Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5228–5235.
- [Haenlein and Kaplan, 2019] Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61, 5–14.
- [Hatuka et al., 2018] Hatuka, T., Rosen-Zvi, I., Birnhack, M., Toch, E., & Zur, H. (2018). The political premises of contemporary urban concepts: The global city, the sustainable city, the resilient city, the creative city, and the smart city. *Planning Theory & Practice*, 19, 160–179.
- [Hou-Liu, 2018] Hou-Liu, J. (2018). *Benchmarking and improving recovery of number of topics in Latent Dirichlet Allocation models*.
- [Jacobi et al., 2016] Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4, 89–106.
- [Jelodar et al., 2019] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., et al. (2019). Latent Dirichlet Allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, 78, 15169–15211.
- [Jeong et al., 2019] Jeong, B., Yoon, J., & Lee, J. M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*, 48, 280–290.
- [Kaur and Bala, 2018] Kaur, J., & Bala, A. (2018). Predicting power for home appliances based on climatic conditions. *International Journal of Energy Sector Management*.
- [Kekäle and Helo, 2014] Kekäle, T., & Helo, P. (2014). The tipping points of technology development. *Quality Innovation Prosperity*, 18, 01–14.
- [Kitchenham et al., 2009] Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51, 7–15.
- [Lee and Cook, 2020] Lee, A. J., & Cook, P. S. (2020). The myth of the “data-driven” society: Exploring the interactions of data interfaces, circulations, and abstractions. *Sociology Compass*, e12749.
- [Lee and Park, 2018] Lee, Y. J., & Park, J. Y. (2018). Identification of future signal based on the quantitative and qualitative text mining: A case study on ethical issues in artificial intelligence. *Quality & Quantity*, 52, 653–667.
- [Maier et al., 2018] Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Nieker, A., Keinert, A., et al. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12, 93–118.
- [Makridakis, 2017] Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46–60.
- [Moher et al., 2009] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6, e1000097.
- [Niwattanakul et al., 2013] Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013). Using of Jaccard coefficient for keywords similarity. *Proceedings of the international multicongference of engineers and computer scientists*, 380–384.
- [Noriega, 2020] Noriega, M. (2020). The application of artificial intelligence in police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, cooperation, and false confessions. *Futures*, 117, 102510.
- [Papadimitriou et al., 1998] Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. (1998). Latent semantic indexing: A probabilistic analysis. *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on principles of database systems*, 159–168. <https://doi.org/10.1145/275487.275505>.
- [Phillips, 2007] Phillips, F. (2007). On s-curves and tipping points. *Technological Forecasting and Social Change*, 74, 715–730.
- [Pons and Latapy, 2005] Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In *International symposium on computer and information sciences* (pp. 284–293).
- [Quinn et al., 2010] Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54, 209–228.
- [Rizzo, 2020] Rizzo, R. (2020). The development of new e-commerce activities in Messina. In *13th international scientific and practical conference-artificial intelligence anthropogenic nature vs. social origin* (pp. 172–183).
- [Roberts et al., 2021] Roberts, H., Cows, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society*, 36, 59–77.
- [Robinson, 2019] Robinson, A. C. (2019). Elements of viral cartography. *Cartography and Geographic Information Science*, 46, 293–310.
- [Robinson, 2020] Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic National Public Policy strategies for artificial intelligence (AI). *Technology in Society*, 63, 101421.
- [Rubin, 2019] Rubin, V. L. (2019). Disinformation and misinformation triangle. *Journal of Documentation*.

- [Schürer et al., 2017] Schürer, O., Stangl, B., Müller, C., & Hubatschke, C. (2017). A first prototype of a space model of cultural meaning by natural-language human-robot interaction. *Proceedings of the 10th international conference on pervasive technologies related to assistive environments*, 339–345.
- [Seel, 2012] Seel, P. B. (2012). *Digital universe: The global telecommunication revolution*. John Wiley & Sons.
- [Shae and Tsai, 2019] Shae, Z., & Tsai, J. (2019). Ai blockchain platform for trusting news. In *2019 IEEE 39th international conference on distributed computing systems (ICDCS)* (pp. 1610–1619).
- [Stover, 2018] Stover, D. (2018). Garlin Gilchrist: Fighting fake news and the information apocalypse. *Bulletin of the Atomic Scientists*, 74, 283–288.
- [Sundnes, 2014] Sundnes, K. O. (2014). Chapter 4: The transactional structure of society: The basic societal functions. *Scandinavian Journal of Public Health*, 36–47.
- [Sunny et al., 2019] Sunny, S., Patrick, L., & Rob, L. (2019). Impact of cultural values on technology acceptance and technology readiness. *International Journal of Hospitality Management*, 77, 89–96.
- [Tight, 2018] Tight, M. (2018). Higher education journals: Their characteristics and contribution. *Higher Education Research & Development*, 37, 607–619.
- [Ting-Toomey and Dorjee, 2018] Ting-Toomey, S., & Dorjee, T. (2018). *Communicating across cultures*. Guilford Publications.
- [Vanderburg, 1985] Vanderburg, W. H. (1985). Technology, society, and culture: A framework for understanding. *Technology in Society*, 7, 411–422.
- [Vesnic-Alujevic et al., 2020] Vesnic-Alujevic, L., Nascimento, S., & Polvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44, 101961.



**Katalin FEHER** is a senior research fellow at Cyber Economy Research Centre and Future of Higher Education at Budapest Business School University of Applied Sciences in Hungary, Fulbright Research Fellow at Drexel University in Philadelphia, US, Bolyai Janos Research Fellow at Hungarian Academy of Sciences, member of editorial board of KOME International Journal of Pure Communication Inquiry (Q1), visiting professor at the Masaryk University in Czechia. Feher's research is focusing on the social-cultural impact of AI and smart technology.



**Attila I. KATONA** is a researcher at the Department of Quantitative Methods, Faculty of Business and Economics, University of Pannonia and a former data scientist from IBM. His research interest includes the treatment of decision risks arising from measurement uncertainty in quality control, and statistical control chart design. His experiences associate with Management Science and Operations Research and Big Data approaches. Attila has valuable publications in Expert Systems with Applications and Computers & Industrial Engineering journals and he is a recipient of the Pro Scientia Golden Medal from the Hungarian Academy of Sciences.